

Normalized Sampling Strategical Random Oversampling towards Improving Classifier Performance of Chemical Composition based Ceramic Sample Classification

M. Shyamala Devi¹, T. Kujani², K. Maithili³ and Kota Venkateswara Rao⁴

¹ Professor, Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India.

Email: shyamalapmr@gmail.com

^{2,3} Assistant Professor, Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India.

⁴ Assistant Professor, Computer Science & Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, India

Email: {kujani@veltech.edu.in, kmaithili@veltech.edu.in, siva278@gmail.com}

Abstract—Clay, earthen elements, powders, and water mixtures are typically manipulated into desired shapes to create ceramics. The ceramic is then dispersed in a blast furnace, which is a high-temperature oven, after it has been molded. Varnishes are paint-like materials that are used to decorate ceramics and are resistant to water. Manually identifying ceramic samples can lead to the discovery of flaws. The chemical composition of the finished ceramic tiles is not visible. Exact identification of ceramic samples is still a difficult task for the user, as it necessitates more hands-on experience. With this overview, this paper focus on designing an effective workflow for classifying the ceramic samples based on chemical composition. The ceramic chemical sample dataset from KAGGLE with the dataset size of 88 with 19 features having 5 classes of ceramic as 0- "FLQ Ceramic", 1- "DYBS Ceramic", 2- "DYNs Ceramic", 3- "DYY Ceramic", 4- "DYMQC Ceramic". The data is preprocessed and normalized for splitting the training and testing dataset. The normalized dataset is fitted to Anova test analysis to extract the significant features. The insignificant features from Anova test is deleted from the original database dataset and anova normalized dataset is fitted to Exhaustive feature selection using KNN, Naive Bayes, Logistic regression, Decision Tree and Random Forest classifier. The exhaustive feature selected dataset is applied to the above-mentioned classifier algorithms to examine the efficiency with the presence and absence of feature scaling. The exhaustive feature selected dataset is again applied to random oversampling in order to improve the classifier performance. The random oversampled exhaustive feature selected dataset is applied to the above mentioned classifier algorithms to compare the performance metrics like precision, accuracy, fscore, recall and run time. Experimental results shows that the decision tree exhibits 49% accuracy with exhaustive feature selected dataset. The same decision tree classifier shows 100% accuracy after applying random oversampling with exhaustive feature selected dataset.

Index Terms— classifier, oversampling, recall, precision, fscore and accuracy.

I. INTRODUCTION

Ceramics are made from materials that are fragile, painful, and amorphous. As a result, they are used to make pots, house brick, tiles, cement, and glass, among other things. Ceramics samples are utilized in a variety of places in environmental turbine engines. Bio-ceramics are used to make dental plates, screws and synthetic joints for bones. Quick and reliable methods for forecasting properties of the material would be immensely beneficial, as accurate estimation of a ceramic material enables this to be aligned to relevant applications. Different modelling methodologies can be used to estimate substance properties based on configurational and computing data, and they can afford quick and accurate results. The paper is organized in which literature review is discussed in section 2 followed by the paper contributions in section 3. The implementation setup and results are discussed in section 4 followed with the conclusion in section 5.

II. LITERATURE REVIEW

The ceramic's formation is governed by the Universal Oil Products benchmark. The goal of this research is to figure out how to get the initial material properties and heating rate to meet the requirements. The design of an immobile earthenware puck is shown at the end of this article [1]. This research aims provide the antiquated ceramic component collection and analysis high quality services manner. The support vector machine is a beneficial different technique for processing data from primitive ceramic tiles. The results obtained of body and gel specimens would be quite suitable if the SVM algorithm is chosen correctly [2]. Ceramics is a substance made from clay that has been shaped and fired. Many ceramic pieces of great diversity have been discovered in Romanian history. Characterization techniques for raw materials and products are used in ceramic materials in culture and heritage [3].

This article suggests an alternative interpretation component categorization technique on inadequate chemical constituents. A fuzzy inference tree is being used to deal with the problem, with terms realised using a classification approach. It derives fuzzy rules out from tree, which serve as a rationalization for the outcome [4]. Coated tiles are made from water-quenched glasses) that are attached to the surface of ecofriendly tiles and then fired. This study provided an in-depth analysis of investigation on windows glazes for use in protective coating and road surface hardwood flooring [5].

This research uses neural networks, which emerge to be among the most essential technique for solving and designing dynamic networks. This mechanism will be evaluated by comparing to the finite difference method to determine which method is the most appropriate for better modelling the behavioral patterns of Polyvinyl chloride disease - causing micro by chicken wings [6]. Because of the structural information requirements, the prediction range of structure-derived descriptors is limited to the registered materials in the crystal structure database. Machine-learning regression based solely on compositional descriptors, on the other hand, is not bound by this constraint, despite the fact that prediction performance may suffer [7].

ANNs are one of "biologically inspired" computational techniques for acquiring complicated, non-linear relationship between the data. These methods have been applied to a broad variety of chemistry problems, which include accurate estimates of the performance of oil-field cements and components with dynamic behavior [8]. In the prediction of ceramic material properties, ANNs have indeed been found to outperform regression techniques [9]. In addition to the conceptual information, this study used topological (bond type, number of occurrences of a structural fragment or functional group) and geometric (moment of inertia, molecular volume, surface area) descriptors as input parameters. There has also been a lot of research done using ANN techniques to predict the electronic characteristics of lead zirconium titanate [10].

III. PROPOSED WORK

The overall architecture of the proposed Random oversampled Exhaustive feature selected work is shown in Fig. 1. Ceramic chemical sample dataset from KAGGLE with the dataset size of 88 with 19 features having 5 classes of ceramic as 0-"FLQ Ceramic", 1-"DYBS Ceramic", 2-"DYNs Ceramic", 3- "DYY Ceramic", 4- "DYMQC Ceramic" is used for this work. The following contributions are provided in this paper.

- (i) The data is preprocessed and normalized for splitting the training and testing dataset.
- (ii) The normalized dataset is fitted to Anova test analysis to extract the significant features. The insignificant features from Anova test is removed from the dataset and anova normalized dataset is fitted to Exhaustive Feature Selection (EFS) using KNN, Logistic regression, Gaussian Naive Bayes, Decision Tree and Random Forest classifier.

- (iii) The exhaustive feature selected dataset is applied to all the classifiers to analyze the performance with the presence and absence of feature scaling.
- (iv) The exhaustive feature selected dataset is again applied to random over-sampling in order to improve the classifier performance.
- (v) The random oversampled exhaustive feature selected dataset is applied to all the classifiers to analyze the performance metrics like precision, recall, accuracy, fscore and run time.

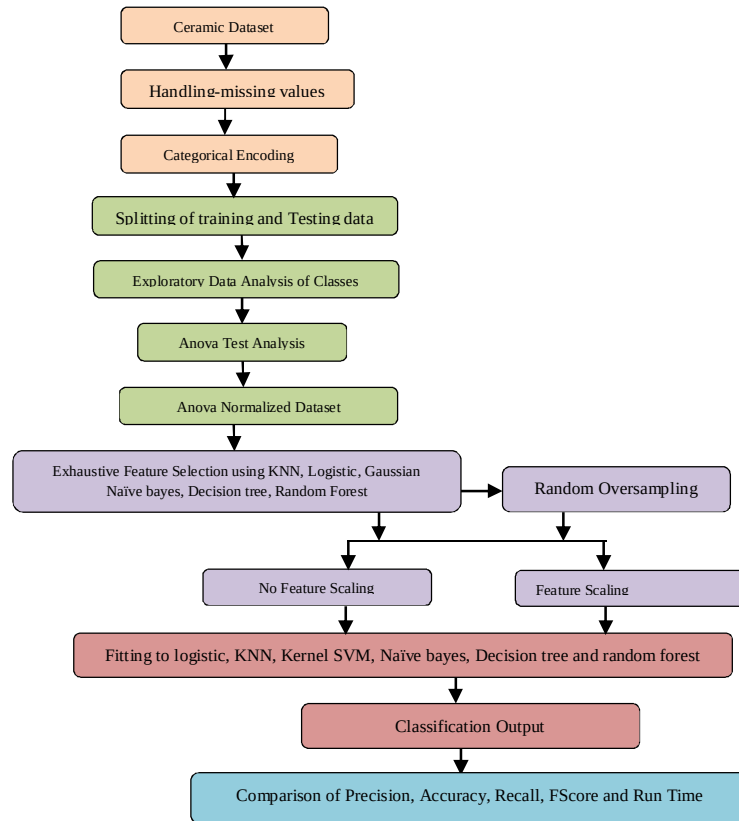


Figure 1. Overall Architecture Flow

IV. RESULTS AND DISCUSSION

A. Implementation Setup

Ceramic chemical sample dataset from KAGGLE with the dataset size of 88 with 19 features having 5 classes of ceramic is used and target distribution is shown in Fig. 2.

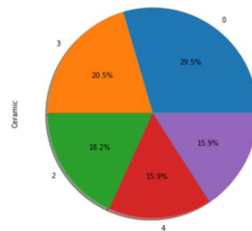


Figure 2. Correlation and Target Distribution

B. Anova Test Analysis

Anova test is used to analyze the features of the dataset by comparing both the null and alternate hypothesis. If the P value associated with the Fstatistic is less than 0.05, then the existence of that feature highly influence the

target. Anova test is applied to dataset features and results shows that the following features {Part, MgO, Al₂O₃, SiO₂, CaO, TiO₂, Fe₂O₃, MnO, CuO, PbO₂, P₂O₅} have value of PR(>F) >0.05 and does not contribute to ceramic sample target as in Table I. The anova normalized dataset is fitted to EFS using KNN, Logistic regression, Gaussian Naive Bayes, Decision Tree and Random Forest classifier and the significant extracted from the EFS is shown in Table II. The EFS from the decision tree classifier shows that least number of best features as ('SrO', 'Y₂O₃').

TABLE I. ANOVA TEST ANALYSIS WITH THE DATASET FEATURES

Features	sum_sq	df	F	PR(>F)
Part	1.3287	1	6.09595	1.0
Na ₂ O	3.2992	1	18.36898	0.000047
MgO	0.0615	1	0.028262	0.866888
Al ₂ O ₃	4.1089	1	1.927346	0.168635
SiO ₂	10.3589	1	5.030448	0.027477
K ₂ O	19.0442	1	9.725091	0.002473
CaO	0.93543	1	0.431308	0.513101
TiO ₂	0.35039	1	0.161055	0.689183
Fe ₂ O ₃	7.83022	1	3.748929	0.056126
MnO	6.51621	1	3.097158	0.081985
CuO	0.30382	1	0.139616	0.709582
ZnO	2.77856	1	1.496577	0.000212
PbO ₂	1.66955	1	0.772841	0.381791
Rb ₂ O	2.28258	1	1.192395	0.000862
SrO	9.01955	1	4.347138	0.040036
Y ₂ O ₃	24.0460	1	12.65515	0.000612
ZrO ₂	4.001171	1	23.41613	0.000006
P ₂ O ₅	1.192806	1	0.550737	0.460039

TABLE II. EXHAUSTIVE FEATURE SELECTION SIGNIFICANT FEATURES

Classifier	Best Subset	Best Features	Accuracy
KNN	(2, 3, 5, 6)	('ZnO', 'Rb ₂ O', 'Y ₂ O ₃ ', 'ZrO ₂ ')	0.54
Logistic	(1, 3, 4, 6)	('K ₂ O', 'Rb ₂ O', 'SrO', 'ZrO ₂ ')	0.51
NaiveBayes	(0, 3, 4, 5)	('Na ₂ O', 'Rb ₂ O', 'SrO', 'Y ₂ O ₃ ')	0.56
Rforest	(4, 5, 6)	('SrO', 'Y ₂ O ₃ ', 'ZrO ₂ ')	0.57
DecisionTree	(4, 5)	('SrO', 'Y ₂ O ₃ ')	0.85

The original dataset is fitted to the classifiers to analyze performance of classifiers and is shown in Table III. The dataset is applied to random oversampling and is fitted to classifiers and is shown in Table IV.

TABLE III. CLASSIFIER PERFORMANCE INDICES BEFORE AND AFTER FEATURE SCALING

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.45	0.50	0.46	0.50	0.18	0.62	0.61	0.56	0.61	0.19
KNN	0.44	0.50	0.42	0.50	0.08	0.74	0.67	0.69	0.67	0.16
KSVM	0.29	0.39	0.27	0.39	1.02	0.37	0.44	0.37	0.44	0.58
GNB	0.55	0.50	0.46	0.50	0.01	0.55	0.50	0.46	0.50	0.00
Dtree	0.35	0.33	0.33	0.33	0.02	0.35	0.33	0.33	0.33	0.01
RFor	0.83	0.72	0.73	0.72	0.09	0.82	0.67	0.66	0.67	0.10

TABLE IV. CLASSIFIER PERFORMANCE OF RANDOM OVERSAMPLED DATASET BEFORE AND AFTER SCALING

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.55	0.53	0.53	0.53	0.04	0.85	0.84	0.84	0.84	0.03
KNN	0.99	0.99	0.99	0.99	0.01	0.98	0.98	0.98	0.98	0.01
KSVM	0.40	0.47	0.42	0.47	0.01	0.97	0.97	0.97	0.97	0.00
GNB	0.67	0.64	0.64	0.64	0.01	0.64	0.60	0.60	0.60	0.00
Dtree	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.00
RFor	0.99	0.99	0.99	0.99	0.16	0.99	0.99	0.99	0.99	0.15

The KNN EFS dataset and KNN EFS random oversampled dataset with features ('ZnO', 'Rb₂O', 'Y₂O₃', 'ZrO₂') is fitted to the classifiers to analyze performance of classifiers and is shown in Table V. and Table VI.

TABLE V. EFS KNN DATASET PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.53	0.50	0.50	0.50	0.04	0.40	0.50	0.43	0.50	0.03
KNN	0.57	0.61	0.58	0.61	0.03	0.67	0.67	0.64	0.67	0.01
KSVM	0.37	0.44	0.37	0.44	0.02	0.71	0.61	0.64	0.61	0.04
GNB	0.52	0.56	0.52	0.56	0.03	0.52	0.56	0.52	0.56	0.03
Dtree	0.53	0.50	0.50	0.50	0.01	0.53	0.50	0.50	0.50	0.05
RFor	0.57	0.56	0.55	0.56	0.13	0.60	0.61	0.60	0.61	0.14

TABLE VI. EFS KNN RANDOM OVERSAMPLED DATASET PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.55	0.53	0.53	0.53	0.04	0.85	0.84	0.84	0.84	0.03
KNN	0.99	0.99	0.99	0.99	0.01	0.98	0.98	0.98	0.98	0.01
KSVM	0.40	0.47	0.42	0.47	0.01	0.97	0.97	0.97	0.97	0.00
GNB	0.67	0.64	0.64	0.64	0.01	0.64	0.60	0.60	0.60	0.00
Dtree	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00	0.00
RFor	0.99	0.99	0.99	0.99	0.16	0.99	0.99	0.99	0.99	0.20

The logistic EFS dataset and logistic EFS random oversampled dataset with features ('K2O', 'Rb2O', 'SrO', 'ZrO2') is fitted to the classifiers to analyze the performance of classifiers and is shown in Table VII. and Table VIII.

TABLE VII. EFS LOGISTIC DATASET PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.55	0.56	0.54	0.56	0.03	0.49	0.61	0.53	0.61	0.00
KNN	0.53	0.56	0.53	0.56	0.00	0.64	0.72	0.67	0.72	0.00
KSVM	0.26	0.50	0.34	0.50	0.02	0.37	0.44	0.38	0.44	0.00
GNB	0.45	0.50	0.45	0.50	0.00	0.45	0.50	0.45	0.50	0.00
Dtree	0.60	0.56	0.56	0.56	0.02	0.60	0.56	0.56	0.56	0.00
RFor	0.47	0.39	0.43	0.39	0.17	0.50	0.44	0.46	0.44	0.16

TABLE VIII. EFS LOGISTIC RANDOM OVERSAMPLED DATASET PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.55	0.53	0.53	0.53	0.05	0.85	0.84	0.84	0.84	0.05
KNN	0.99	0.99	0.99	0.99	0.02	0.98	0.98	0.98	0.98	0.02
KSVM	0.40	0.47	0.42	0.47	0.02	0.97	0.97	0.97	0.97	0.02
GNB	0.67	0.64	0.64	0.64	0.00	0.64	0.60	0.60	0.60	0.00
Dtree	1.00	1.00	1.00	1.00	0.01	1.00	1.00	1.00	1.00	0.01
RFor	0.99	0.99	0.99	0.99	0.21	0.99	0.99	0.99	0.99	0.21

The Naive bayes EFS dataset and Naive bayes EFS random oversampled dataset with features ('Na2O', 'Rb2O', 'SrO', 'Y2O3') is fitted to the classifiers to analyze performance of classifiers and is shown in Table IX. and Table X.

TABLE IX. EFS NAIVE BAYES DATASET PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.59	0.56	0.53	0.56	0.02	0.59	0.61	0.56	0.61	0.00
KNN	0.59	0.56	0.56	0.56	0.00	0.67	0.61	0.60	0.61	0.00
KSVM	0.18	0.39	0.25	0.39	0.00	0.64	0.61	0.57	0.61	0.00
GNB	0.42	0.50	0.45	0.50	0.00	0.42	0.50	0.45	0.50	0.00
Dtree	0.65	0.61	0.63	0.61	0.00	0.65	0.61	0.63	0.61	0.00
RFor	0.60	0.61	0.59	0.61	0.12	0.60	0.61	0.59	0.61	0.15

TABLE X. EFS EFS NAIVE BAYES RANDOM OVERSAMPLED PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.55	0.53	0.53	0.53	0.03	0.85	0.84	0.84	0.84	0.05
KNN	0.99	0.99	0.99	0.99	0.02	0.98	0.98	0.98	0.98	0.02
KSVM	0.40	0.47	0.42	0.47	0.02	0.97	0.97	0.97	0.97	0.02
GNB	0.67	0.64	0.64	0.64	0.03	0.64	0.60	0.60	0.60	0.00
Dtree	1.00	1.00	1.00	1.00	0.02	1.00	1.00	1.00	1.00	0.01
RFor	0.99	0.99	0.99	0.99	0.21	0.99	0.99	0.99	0.99	0.21

The Random Forest EFS dataset and Random Forest EFS random oversampled dataset with features ('SrO', 'Y2O3', 'ZrO2') is fitted to the classifiers to analyze the performance of classifiers and is shown in Table X1. and Table XII.

TABLE XI. EFS RANDOM FOREST DATASET PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.44	0.44	0.40	0.44	0.02	0.62	0.56	0.54	0.56	0.00
KNN	0.65	0.67	0.65	0.67	0.01	0.57	0.67	0.61	0.67	0.00
KSVM	0.26	0.50	0.34	0.50	0.00	0.50	0.61	0.53	0.61	0.02
GNB	0.60	0.61	0.58	0.61	0.00	0.60	0.61	0.58	0.61	0.00
Dtree	0.59	0.50	0.51	0.50	0.00	0.59	0.50	0.51	0.50	0.00
RFor	0.75	0.72	0.72	0.72	0.12	0.75	0.72	0.72	0.72	0.15

TABLE XII. EFS RANDOM FOREST RANDOM OVERSAMPLED PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.55	0.53	0.53	0.53	0.02	0.85	0.84	0.84	0.84	0.03
KNN	0.99	0.99	0.99	0.99	0.05	0.98	0.98	0.98	0.98	0.02
KSVM	0.40	0.47	0.42	0.47	0.01	0.97	0.97	0.97	0.97	0.02
GNB	0.67	0.64	0.64	0.64	0.06	0.64	0.60	0.60	0.60	0.01
Dtree	1.00	1.00	1.00	1.00	0.02	1.00	1.00	1.00	1.00	0.03
RFor	0.99	0.99	0.99	0.99	0.12	0.99	0.99	0.99	0.99	0.18

The Decision Tree EFS dataset and Decision Tree EFS random oversampled dataset with features ('SrO', 'Y2O3') is fitted to the classifiers to analyze the performance of classifiers and is shown in Table XIII. and Table XIV.

TABLE XIII. EFS DECISION TREE DATASET PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.50	0.56	0.47	0.56	0.02	0.35	0.50	0.40	0.50	0.01
KNN	0.38	0.44	0.41	0.44	0.00	0.49	0.56	0.52	0.56	0.00
KSVM	0.18	0.39	0.25	0.39	0.00	0.46	0.61	0.51	0.61	0.00
GNB	0.63	0.61	0.53	0.61	0.00	0.63	0.61	0.53	0.61	0.00
Dtree	0.57	0.56	0.54	0.56	0.00	0.57	0.56	0.54	0.56	0.00
RFor	0.61	0.61	0.61	0.61	0.12	0.61	0.61	0.61	0.61	0.14

TABLE XIV. EFS DECISION TREE RANDOM OVERSAMPLED PERFORMANCE BEFORE AND AFTER SCALING.

Classifier	Before Feature Scaling					After Feature Scaling				
	Precision	Rcall	FScore	Accu	RunTime	Precision	Recall	FScore	Accu	RunTime
LReg	0.55	0.53	0.53	0.53	0.05	0.85	0.84	0.84	0.84	0.03
KNN	0.99	0.99	0.99	0.99	0.04	0.98	0.98	0.98	0.98	0.02
KSVM	0.40	0.47	0.42	0.47	0.03	0.97	0.97	0.97	0.97	0.02
GNB	0.67	0.64	0.64	0.64	0.06	0.64	0.60	0.60	0.60	0.01
Dtree	1.00	1.00	1.00	1.00	0.02	1.00	1.00	1.00	1.00	0.03
RFor	0.99	0.99	0.99	0.99	0.12	0.99	0.99	0.99	0.99	0.17

V. CONCLUSIONS

This paper attempts to analyse the performance of the original dataset, oversampled dataset, exhaustive feature selected dataset, random oversampled exhaustive feature selected dataset of the ceramic chemical sample dataset while fitting to the classifiers. This paper is implemented to prove how well the exhaustive feature selected oversampled decision tree dataset with ('SrO', 'Y2O3') is performing while fitting to the classifiers. Experimental results shows that the decision tree classifier exhibits 49% accuracy with exhaustive feature selected dataset. The same decision tree classifier shows 100% accuracy after applying random oversampling with exhaustive feature selected dataset.

REFERENCES

- [1] A. G. Edy Sutjipto, A. Hisyam, N. Salim, M. H. Ab Rahim, A. Legowo, and M. H. Ani, "Development of inert ceramic for industrial application based on ternary phase diagram of potassium oxide-aluminum oxide-silicon dioxide," *Advances in Science and Engineering Technology*, pp. 1-4, 2019.
- [2] L. H. Yang, and H. Huang, "A Classification Method of Ancient Ceramics Based on Support Vector Machine in Ceramic Cloud Service Platform," *In the Proceedings of the International Conference on Big Data Security on Cloud*, pp. 108-112, 2019.
- [3] J. Gasteiger, and J. Zupan, "Neural Networks in Chemistry," *Angewandte Chemie International Edition*, vol. 32, no. 4, pp. 503-527, 1993.
- [4] Arnaud Grivet Sebert, and Jean-Philippe Poli, "Material Classification from Imprecise Chemical Composition : Probabilistic vs Possibilistic Approach," *In the Proceedings of the IEEE International Conference on Fuzzy Systems*, 2018.
- [5] Rincon, Jesus, Romero, and Maximina, "Glass-Ceramic Glazes for Ceramic Tiles: A Review," *Journal of Materials Science*, vol. 47, pp. 553-582, 2012.
- [6] Lakhdar, Abdelghani Moumen, Aziz Laabid, Zineb Mansouri, and Khalifa, "Prediction of the Behavior of Bio-Loaded Flexible PVC by Chicken Feathers with Artificial Neural Networks," *Web of Conferences*, 2022.
- [7] Taruto Atsumi, Kosei Sato, Yudai Yamaguchi, Masato Hamaie, Risa Yasuda, Naoto Tanibata, Hayami Takeda, Masanobu Nakayama, Masayuki Karasuyama, and Ichiro Takeuchi, "Chemical Composition Data-Driven Machine-Learning Prediction for Phase Stability and Materials Properties of Inorganic Crystalline Solids," *Basic solid state physics*, 2022.
- [8] P. V. Coveney, and W. Humphries, "Molecular modelling of the mechanism of action of phosphonate retarders on hydrating cements," *Journal of the Chemical Society, Faraday Transactions*, vol. 92, pp. 831-841, 1996.
- [9] D. Guo, L. Li, C. Nan, J. Xia, and Z. Gui, "Modeling and analysis of the electrical properties of PZT through neural networks," *Journal of the European Ceramic Society*, vol. 23, pp. 2177-2181, 2003.
- [10] K. Cai, J. Xia, and L. Z. Gui, "Analysis of the electrical properties of PZT by a BP artificial neural network," *Computational Materials Science*, vol. 34, pp. 166-172, 2005.